

Sequence Listing

SEQ ID NO:1

the major 17 amino acid repeat,

GluGlnGlnSerAspLeuGluGlnGluArgLeuAlaLysGluLysLeuGln (SEQ ID NO:1)

SEQ ID NO:2

minor repeat

GluGlnGlnArgAspLeuGluGlnGluArgLeuAlaLysGluLysLeuGln (SEQ ID NO:2)

SEQ ID NO:3

DNA sequence of the gene *lsa-nrc*^{Hmut} Arg mutant

```

1  ATGGGTACCA ACAGCGAAAA AGACGAAATT ATCAAAAGCA ATCTCCGCTC CGGCAGCTCC
61  AACAGCCGCA ACCGCATCAA CGAGGAAAAG CATGAGAAGA AACATGTGCT GAGCCACAAC
121 TCCTACGAGA AGACTAAAAA CAACGAAAAC AACAAATTCT TTGACAAGGA CAAAGAGCTG
181 ACGATGAGCA ACGTTAAAAA CGTATCCCAG ACCAACTTTA AATCCCTCCT GCGCAACCTC
241 GCGGTTTCCG AGAACATCTT TCTCAAAGAA AACAACTGA ACAAGGAAGG CAAACTGATT
301 GAACATATCA TCAACGACGA CGATGACAAA AAAAAATACA TTAAAGGCCA GGATGAAAAAT
361 CGCCAGGAAG ACCTCGAAGA AAAAGCTGCT GAACAGCAGT CGGACCTGGA ACAGGAGCGC
421 CTCGCTAAAG AAAAGCTCCA GGAGCGCCTC GCTAAAGAAA AGCTCCAGGA GCAACAGCGC
481 GACCTGGAAC AGCGCAAGGC TGACACGAAA AAAACCTGG AACGCAAAAA GGAACACGCG
541 GACGTTCTGG CTGAGGACCT GTACGGCCGC CTGGAAATCC CAGCTATCGA ACTCCCATCC
601 GAAAACGAAC GCGGCTACTA CATCCCACAC CAGAGCAGCC TGCCACAAGA TAATCGCGGG
661 AACTCCCGCG ACAGTAAGGA AATCAGCATC ATCGAAAAAA CCAACCGCGA AAGCATTACC
721 ACCAACGTGG AAGGCCGCCG CGACATCCAC AAAGGCCACC TCGAAGAAAA GAAAGACGGC
781 TCCATCAAAC CAGAACAGAA AGAAGACAAA AGCGCTGATA TCCAGAACCA CACCCTGGAG
841 ACCGTGAACA TTAGCGACGT GAACGACTTC CAGATCAGCA AGTACGAGGA CGAAATCTCC
901 GCTGAATACG ATGACTCCCT GATCGACGAA GAAGAAGACG ACGAAGATCT GGATGAATTC
961 AAACCAATTG TCCAGTACGA TAACTTTCAG GACGAAGAAA ATATCGGCAT TTACAAAGAA
1021 CTCGAAGACC TCATCGAGAA AAACGAAAAC CTGGACGACC TGGACGAAGG CATCGAAAAA
1081 TCCTCCGAAG AACTGAGCGA AGAAAAAATC AAAAAAGGCA AGAAATACGA AAAAACCAG
1141 GACAACAAC TCAAACCAA CGACAAATCC CTCTACGACG AGCACATTAA AAAATACAA
1201 AACGACAAGC AAGTGAACAA GGAAAAGGAA AAATTTATCA AATCCCTCTT CCACATCTTC
1261 GATGGCGATA ACGAAATTCT GCAAATTGTA GACGAACGGT TGAGCGAAGA CATCACTAAA
1321 TACTTCATGA AGCTTGGGGG CTCCGGTTCT CCACACCACC ACCACCACCA CTGA

```

SEQ ID NO:4

LSA-NRC(H) Mut Protein

MetGlyThrAsnSerGluLysAspGluIleIleLysSerAsnLeuArgS
erGlySerSerAsnSerArgAsnArg 25

IleAsnGluGluLysHisGluLysLysHisValLeuSerHisAsnSerT
yrGluLysThrLysAsnAsnGluAsn 50

AsnLysPhePheAspLysAspLysGluLeuThrMetSerAsnValLysA
snValSerGlnThrAsnPheLysSer 75

LeuLeuArgAsnLeuGlyValSerGluAsnIlePheLeuLysGluAsnL
ysLeuAsnLysGluGlyLysLeuIle 100

GluHisIleIleAsnAspAspAspAspLysLysLysTyrIleLysGlyG
lnAspGluAsnArgGlnGluAspLeu 125

GluGluLysAlaAlaGluGlnGlnSerAspLeuGluGlnGluArgLeuA
laLysGluLysLeuGlnGluArgLeu 150

AlaLysGluLysLeuGlnGluGlnGlnArgAspLeuGluGlnArgLysA
laAspThrLysLysAsnLeuGluArg 175

LysLysGluHisGlyAspValLeuAlaGluAspLeuTyrGlyArgLeuG
luIleProAlaIleGluLeuProSer 200

GluAsnGluArgGlyTyrTyrIleProHisGlnSerSerLeuProGlnA
spAsnArgGlyAsnSerArgAspSer 225

LysGluIleSerIleIleGluLysThrAsnArgGluSerIleThrThrA
snValGluGlyArgArgAspIleHis 250

LysGlyHisLeuGluGluLysLysAspGlySerIleLysProGluGlnL
ysGluAspLysSerAlaAspIleGln 275

AsnHisThrLeuGluThrValAsnIleSerAspValAsnAspPheGlnI
leSerLysTyrGluAspGluIleSer 300

AlaGluTyrAspAspSerLeuIleAspGluGluGluAspAspGluAspL
euAspGluPheLysProIleValGln 325

TyrAspAsnPheGlnAspGluGluAsnIleGlyIleTyrLysGluLeuG
luAspLeuIleGluLysAsnGluAsn 350

LeuAspAspLeuAspGluGlyIleGluLysSerSerGluGluLeuSerG
luGluLysIleLysLysGlyLysLys 375

TyrGluLysThrLysAspAsnAsnPheLysProAsnAspLysSerLeuT
yrAspGluHisIleLysLysTyrLys 400

AsnAspLysGlnValAsnLysGluLysGluLysPheIleLysSerLeuP
heHisIlePheAspGlyAspAsnGlu 425

IleLeuGlnIleValAspGlu**Arg**LeuSerGluAspIleThrLysTyrP
heMetLysLeuGlyGlySerGlySerPro 450

HisHisHisHisHisHis 456

SEQ ID NO:5

conserved amino acids of the same basic 17 amino acids following the order:

X₁GlnGlnX₂AspX₃GluGlnX₄ArgX₅AlaX₆GluX₇LeuGln (SEQ ID NO:5) where x₁ is either Glu or Gly; x₂ is Ser or Arg; x₃ is Asp or Ser; x₄ is Glu or Asp; x₅ is Leu or Arg; x₆ is Lys or Asn and x₇ is Lys or Thr or Arg.

SEQ ID NO:6

LeuThrMetSerAsnValLysAsnValSerGlnThrAsnPheLysSerLeuLeuArgAsnLeuGlyValSer
(SEQ ID NO:6)

SEQ ID NO:7

GluGlnGlnSerAspLeuGluGlnGluArgLeuAlaLysGluLysLeuGln (SEQ ID NO:7)

SEQ ID NO:8

GluArgLeuAlaLysGluLysLeuGlnGluGlnGlnArgAspLeuGluGln (SEQ ID NO:8)

SEQ ID NO:9

ThrLysLysAsnLeuGluArgLysLysGluHisGlyAspValLeuAlaGluAspLeuTyr (SEQ ID NO:9)

SEQ ID NO:10

AsnSerArgAspSerLysGluIleSerIleIleGluLysThrAsnArgGluSerIleThrThrAsnValGlu
GlyArgArgAspIleHisLysGlyHisLeu (SEQ ID NO:10)

SEQ ID NO:11

LysProIleValGlnTyrAspAsnPhe (SEQ ID NO:11)

SEQ ID NO:12

AsnGluAsnLeuAspAspLeuAspGluGlyIleGluLysSerSerGluGluLeuSerGluGluLys
Ile (SEQ ID NO:12)

SEQ ID NO:13

LysProAsnAspLysSerLeu (SEQ ID NO:13)

SEQ ID NO:14

AspAsnGluIleLeuGlnIleValAspGluLeuSerGluAspIleThrLysTyrPheMetLysLeu
(SEQ ID NO:14)

SEQ ID NO:15

AspAsnGluIleLeuGlnIleValAspGlu**Arg**LeuSerGluAspIleThrLysTyrPheMetLysLeu
(SEQ ID NO:15)

SEQ ID NO:16

LeuThrMetSerAsnValLysAsnValSerGlnThrAsnPheLysSerLeuLeuArgAsnLeuGlyValSer
(SEQ ID NO:16)

SEQ ID NO:17

HisThrLeuGluThrValAsnIleSerAspValAsnAspPheGlnIleSerLysTyrGlu (SEQ ID NO:17)

SEQ ID NO:18

AspGluAspLeuAspGluPheLysProIleValGlnTyrAspAsnPheGlnAsp (SEQ ID NO:18)

SEQ ID NO:19

IleGlyIleTyrLysGluLeuGluAspLeuIleGluLys (SEQ ID NO:19)

SEQ ID NO:20

AsnGluAsnLeuAspAspLeuAspGluGlyIleGluLysSerSerGluGluLeuSerGluGluLysIle
(SEQ ID NO:20)

SEQ ID NO:21

IleLysLysGlyLysLysTyrGluLysThrLysAspAsnAsnPhe (SEQ ID NO:21)

SEQ ID NO:22

AspAsnGluIleLeuGlnIleValAspGluLeuSerGluAspIleThrLysTyrPheMetLysLeu (SEQ ID NO:22)

SEQ ID NO:23

TyrTyrIleProHisGlnSerSerLeu (SEQ ID NO:23)

SEQ ID NO:24

Amino acid sequence of LSA-NRC(H) repeat sequence between N & C terminals.

GluGlnGlnSerAspLeuGluGlnGluArgLeuAlaLysGluLysLeuGln

GluArgLeuAlaLysGluLysLeuGlnGluGlnGlnArgAspLeuGluGln (SEQ ID NO:24).

SEQ ID NO:25

DNA sequence of the gene *lsa-nrc^H*

```

1  ATGGGTACCA ACAGCGAAAA AGACGAAATT ATCAAAAGCA ATCTCCGCTC CGGCAGCTCC
61  AACAGCCGCA ACCGCATCAA CGAGGAAAAG CATGAGAAGA AACATGTGCT GAGCCACAAC
121 TCCTACGAGA AGACTAAAAA CAACGAAAAC AACAAATTCT TTGACAAGGA CAAAGAGCTG
181 ACGATGAGCA ACGTTAAAAA CGTATCCCAG ACCAACTTTA AATCCCTCCT GCGCAACCTC
241 GGCGTTTCCG AGAACATCTT TCTCAAAGAA AACAACTGA ACAAGGAAGG CAAACTGATT
301 GAACATATCA TCAACGACGA CGATGACAAA AAAAAATACA TTAAAGGCCA GGATGAAAAT
361 CGCCAGGAAG ACCTCGAAGA AAAAGCTGCT GAACAGCAGT CGGACCTGGA ACAGGAGCGC
421 CTCGCTAAAG AAAAGCTCCA GGAGCGCCTC GCTAAAGAAA AGCTCCAGGA GCAACAGCGC
481 GACCTGGAAC AGCGCAAGGC TGACACGAAA AAAACCTGG AACGCAAAAA GGAACACGGC

```

541 GACGTTCTGG CTGAGGACCT GTACGGCCGC CTGGAAATCC CAGCTATCGA ACTCCCATCC
 601 GAAAACGAAC GCGGCTACTA CATCCCACAC CAGAGCAGCC TGCCACAAGA TAATCGCGGG
 661 AACTCCCGCG ACAGTAAGGA AATCAGCATC ATCGAAAAAA CCAACCGCGA AAGCATTACC
 721 ACCAACGTGG AAGGCCGCG CGACATCCAC AAAGGCCACC TCGAAGAAAA GAAAGACGGC
 781 TCCATCAAAC CAGAACAGAA AGAAGACAAA AGCGCTGATA TCCAGAACCA CACCCTGGAG
 841 ACCGTGAACA TTAGCGACGT GAACGACTTC CAGATCAGCA AGTACGAGGA CGAAATCTCC
 901 GCTGAATACG ATGACTCCCT GATCGACGAA GAAGAAGACG ACGAAGATCT GGATGAATTC
 961 AAACCAATTG TCCAGTACGA TAACTTTCAG GACGAAGAAA ATATCGGCAT TTACAAAGAA
 1021 CTCGAAGACC TCATCGAGAA AAACGAAAAC CTGGACGACC TGGACGAAGG CATCGAAAAA
 1081 TCCTCCGAAG AACTGAGCGA AGAAAAATC AAAAAAGGCA AGAAATACGA AAAAACCAG
 1141 GACAACAAC TCAAACCAA CGACAAATCC CTCTACGACG AGCACATTAA AAAATACAAA
 1201 AACGACAAGC AAGTGAACAA GGAAAAGGAA AAATTTATCA AATCCCTCTT CCACATCTTC
 1261 GATGGCGATA ACGAAATTCT GCAAATTGTA GACGAACTGA GCGAAGACAT CACTAAATAC
 1321 TTCATGAAGC TTGGGGGCTC CGGTTCTCCA CACCACCACC ACCACCACTG A

SEQ ID NO:26

LSA-NRC Protein

MetGlyThrAsnSerGluLysAspGluIleIleLysSerAsnLeuArgS
erGlySerSerAsnSerArgAsnArg 25

IleAsnGluGluLysHisGluLysLysHisValLeuSerHisAsnSerT
yrGluLysThrLysAsnAsnGluAsn 50

AsnLysPhePheAspLysAspLysGluLeuThrMetSerAsnValLysA
snValSerGlnThrAsnPheLysSer 75

LeuLeuArgAsnLeuGlyValSerGluAsnIlePheLeuLysGluAsnL
ysLeuAsnLysGluGlyLysLeuIle 100

GluHisIleIleAsnAspAspAspLysLysLysTyrIleLysGlyG
lnAspGluAsnArgGlnGluAspLeu 125

GluGluLysAlaAlaGluGlnGlnSerAspLeuGluGlnGluArgLeuA
laLysGluLysLeuGlnGluArgLeu 150

AlaLysGluLysLeuGlnGluGlnGlnArgAspLeuGluGlnArgLysA
laAspThrLysLysAsnLeuGluArg 175

LysLysGluHisGlyAspValLeuAlaGluAspLeuTyrGlyArgLeuG
luIleProAlaIleGluLeuProSer 200

GluAsnGluArgGlyTyrTyrIleProHisGlnSerSerLeuProGlnA
spAsnArgGlyAsnSerArgAspSer 225

LysGluIleSerIleIleGluLysThrAsnArgGluSerIleThrThrA
snValGluGlyArgArgAspIleHis 250

LysGlyHisLeuGluGluLysLysAspGlySerIleLysProGluGlnL
ysGluAspLysSerAlaAspIleGln 275

AsnHisThrLeuGluThrValAsnIleSerAspValAsnAspPheGlnI
leSerLysTyrGluAspGluIleSer 300

AlaGluTyrAspAspSerLeuIleAspGluGluGluAspAspGluAspL
euAspGluPheLysProIleValGln 325

TyrAspAsnPheGlnAspGluGluAsnIleGlyIleTyrLysGluLeuG
luAspLeuIleGluLysAsnGluAsn 350

LeuAspAspLeuAspGluGlyIleGluLysSerSerGluGluLeuSerG
luGluLysIleLysLysGlyLysLys 375

TyrGluLysThrLysAspAsnAsnPheLysProAsnAspLysSerLeuT
yrAspGluHisIleLysLysTyrLys 400

AsnAspLysGlnValAsnLysGluLysGluLysPheIleLysSerLeuP
heHisIlePheAspGlyAspAsnGlu 425

IleLeuGlnIleValAspGluLeuSerGluAspIleThrLysTyrPheM
etLysLeuGlyGlySerGlySerPro 450

HisHisHisHisHisHis 456

SEQ ID NO:27

VSQTNFKSL (SEQ ID NO:27)

SEQ ID NO:28

SQTNFKSL (SEQ ID NO:28),